A Closer Look at the Quality of Small Area Estimates from the American Community Survey

Susan P. Love and Deborah H. Griffin
U.S. Bureau of the Census
Washington D.C. 20233

Presented at the 2003 Joint Statistical Meetings in San Francisco, CA
August 3 - 7, 2003

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The authors acknowledge significant contributions from Don Dalzell, Katie Bench, Michael Starsinic, and John Stiller

Key Words: self-response, unit nonresponse, item nonresponse, nonsampling error

Introduction

Current plans call for the American Community Survey (ACS) to replace the decennial census sample as the premier source of detailed demographic and socio-economic information about the nation's population and housing. In doing so it takes on an incredibly important role. It must produce reasonable, accurate, reliable, and consistent data for every area of the country from the national and state level to the tract and block group level, and do so year after year. By providing data annually, the bar has been raised for the ACS above that ever required for the census sample, which had to produce these data only once every ten years. It is the ACS data for small areas, particularly for areas that are less densely populated, that has been the concern of rural demographers, researchers, planners, and decision-makers who in the past have had to rely on the decennial census sample estimates refreshed only once a decade and reflecting a single point in time. Until the introduction of the ACS, no single current survey was producing estimates of the wide-ranging characteristics covered by the census sample, and none had samples of a size sufficient to provide statistically reliable estimates for areas with populations below 100,000. The ACS was designed specifically to build and improve upon the solid tradition of both the census sample and the current surveys.

This paper looks at the ability of the ACS to produce quality data at the county level by investigating several aspects of standard survey methods and focusing mainly on nonsampling error issues. We have confined this analysis to the 21 ACS test sites with populations of at least 65,000, the sites for which yearly survey estimates equivalent to those derived from the Census 2000 sample have been released. The sites consist of 24 of the 36 counties that have been continuously sampled and surveyed under the ACS systematic design using ACS data collection methods.

The Evolution of the Decennial Census Sample [1]

The ACS is replacing a national sample survey that has evolved over many decades. The decennial census sample has been in existence for seven censuses. Probability sampling was introduced slowly into the census as an innovative application of statistical techniques. In the 1940 census, sampling was used to allow several additional questions to be asked of 5 percent of the enumerated population without a noticeable increase in respondent burden and more cheaply than the cost of adding these "sample" questions to the full census data collection effort. The estimates were made simply by weighting the items by 20. In 1950 the "sample" population was increased to 20 percent, and the sampling concept was extended to the housing data collection, but the two were not linked. Again, estimation was accomplished by adding up the answers that had been multiplied by the reciprocal of the sampling fraction, or 5. In 1960 the sampling was changed so that the housing unit was the primary sampling unit, and all people who occupied a sample unit formed the population sample. The overall expected size of the sample was increased to 25 percent, and two housing unit samples were taken - a 20 percent and a 5 percent sample – each collecting different information. Data from the samples were provided for areas as small as tracts, which average 4000 people. The more extensive use of sampling and its release for small areas introduced moderate amounts of sampling error into the estimates. In an attempt to control the variance, ratio estimation of the sample data to the full census counts was introduced instead of the simple weighting by probabilities of selection used previously. The unweighted count of units in the 1960 census sample was 24.5 percent of the total housing count.

The 1970 census sampling rate was decreased to 20 percent, made up of two smaller interpenetrating samples of 15 percent and 5 percent. Three different

[1] 1960 Census of Population and Housing Procedural History; Measuring America: The Decennial Censuses from 1790 to 2000

housing unit forms were used – one regular census form ("short") and two sample forms that included the items on the regular form and shared a subset of the sample items. The 1960 census had the largest sampling rate but the 1970 census had the most complicated sample design. The unweighted count of units in the 1970 census sample was 19.6 percent of the total housing count.

In 1980 the sampling rate was decreased again, and differential sampling was introduced. Two sampling rates were used – ½ and 1/6 – and the unweighted count of units placed in the census sample was 18.4 percent of the total housing count. A third rate of 1/8 was introduced in the 1990 census. The actual sample consisted of 16.0 percent of the total housing units, while the expected sample size was 17.6 percent. A fourth sampling rate of 1/4 was added in 2000 and 15.8 percent of housing units were actually placed in the 2000 sample–a sample that was expected to contain 17.3 percent . The ever-increasing number of housing units in the nation, the gap between the expected and the realized census sample size, and the fact that enumerating the population is becoming more difficult and costly every decade certainly points to having to make a more radical change in 2010 than merely adding another sampling rate.

The census sample has always been a work in progress. Every decennial census has been a unique entity, its questionnaires, data collection procedures, and even its sampling and estimation methods changing every ten years. The ACS is the next chapter of this 70-year saga of census samples. It is a major innovative step to meeting the nation's need for the kind of information that has only been available through the modern census samples. By adopting the concept of continuous measurement and spreading a sample of 3 million housing units every year over twelve months, and by using the best mail survey techniques combined with computer-assisted technology and a permanent interviewing staff, the ACS will produce estimates of the demographic and socio-economic characteristics of the nation updated annually.

The Treatment of Nonresponse in the Census Sample and the ACS

Much has been written about the relationship between sample size and sampling error and how the ACS sample aggregated over 5 years will not be as large as past decennial samples. It is relatively straightforward, using standard formulas for calculating variances, to arrive at a conclusion that estimates produced by the ACS will be less precise than those that have been produced every ten years by the census sample. The Census Bureau estimates that coefficients of variation for the ACS five-year averages will be about 1.33 times as large as the coefficients of variation for Census 2000 sample estimates.[2] But is it that simple? There is no doubt that estimates and distributions produced by the ACS will be based on samples smaller than past census samples. However, judgments concerning the accuracy of estimates should also take into account levels of nonsampling error. This paper will focus on the nonresponse or noninterview aspects of nonsampling error, the kind of error that reduces the effective size of samples, thereby increasing the variance on the estimates, and also introduces the potential for biases in the data.

Any comparison of the ACS and decennial census sample estimates is incomplete without a discussion of how sample nonresponse is dealt with by each program. The census sample, over its seven-census history, has been troubled with potential biases. Originally, bias was introduced into the data because the sample selection was not controlled. Enumerators could decide which people, and later, which housing units would be selected within their assignment areas, and their decisions were often based on the characteristics of the units themselves. The high levels of enumerator bias in early census samples provided the impetus for introducing self-response methods of data collection into the census. Ironically, the increasing disparity in recent censuses between the completeness of census data for self-response households and households enumerated through followup operations may have introduced a new bias.

Since the advent of computers (circa 1960), for units to be considered members of the census sample, the questionnaires enumerating them had to contain a minimum amount of data. Criteria were devised that identified long form enumerations with no valid sample information, and the units they represented were removed from the census sample processing, thereby decreasing the effective sample size. This has not been a census estimation concern because the units placed in the census sample are weighted to agree with the full "100%" census counts.

Concern about the robustness of the census sample edits and their ability to impute entire sets of sample data for potentially large numbers of households in small areas also played a part in the treatment of nonresponse. The census content edits (and the ACS content edits as well) use a "nearest neighbor hot deck" approach to item imputation. It has been shown that donor matrices used for editing and imputing items in the census can be defined in ways that maintain homogeneity between donors and item nonrespondents, thereby producing unbiased imputations as long as the items are "missed at random." With high item response rates, any residual biases due to this assumption are

[2] Meeting 21st Century Demographic Data Needs –Implementing the American Community Survey: Demonstrating Survey Quality," U.S. Census Bureau, May 2002

small.[3] However, there was a reluctance to place units with no sample data in the sample editing process and possibly jeopardize the ability of the edit to impute reasonable item values, since every single characteristic for these units would have to be obtained from donors. Regardless, the detailed demographic and socio-economic characteristics of these noninterview units are not represented in the census sample.

While nearly all mail return long form questionnaires meet the data requirements for the census sample, considerable numbers of long forms completed by census enumerators do not. The census sample estimation process does not involve special weighting adjustments for noninterview units, and to the extent that their characteristics differ in important ways from the units that meet the census sample requirements, nonresponse bias is introduced. The census sample has a tendency to over-represent the characteristics of the self-responding households and to under-represent the households enumerated in follow-up operations. This disparity can vary from area to area and depends not only on the willingness and ability of residents to understand, fill out, and return the census long form questionnaire, but also on the knowledge and dedication of the temporary enumerators tasked with collecting decennial sample data.

The ACS uses typical current survey methods to deal with sample nonresponse. In an attempt to control the amount of nonresponse bias that is introduced into the estimates, sample units that are not successfully interviewed are adjusted for by a series of weighting factors in the estimation process that take into account both geography and mode of collection. We will discuss ACS and census sample nonresponse in more detail when we compare actual levels, but one more point should be made. The use of subsampling for nonresponse in the final personal interviewing stage of ACS data collection is of particular concern to people who depend on small area data. This is not an issue of nonresponse but of sampling error, it is a legitimate concern, and research is planned to assess the impact of this subsampling on important survey estimates. Units are removed from the sample, but not because they were not successfully interviewed. A systematic probability sample of about one-third of the ACS units that did not respond by either mail or through the computer-assisted telephone operation is selected for personal visit interviewing. This subsampling is done strictly to limit cost, and it results in larger variances on survey estimates, especially of characteristics of the population and housing cohorts heavily represented in the final data collection stage[4]. The ACS subsampling increases the sampling error on the survey estimates, but it does not introduce a potential bias into the overall results.

Measuring the Quality of the ACS Data for Small Areas

The ACS is a hybrid survey. It combines the best decennial self-response methods with the best current survey follow-up methods. It was designed and implemented to continuously and consistently produce complete and accurate demographic and socio-economic data for all types of areas. The development and testing of the ACS design and methodology began in earnest in 1994, and the first mail-out of the ACS questionnaire package to the four initial test sites occurred in October 1995. By 1999, the number of test areas had expanded to 36 counties spread across the country. Survey data have been publicly available for 21 of the test sites consisting of 24 of these counties since 1999, and these data can be accessed through the Census Bureau's website. The 21 sites represent areas with populations that meet the 65,000 minimum required for yearly data-release and are the focus of this paper. Although in many instances the sites are not "small" in terms of population, they are a diverse set of areas that have been continuously sampled and surveyed under the ACS design and methods, and therefore are the primary source of information on the relative performance of the ACS and Census 2000 in a wide variety of areas. Selected site characteristics are shown in Table 1. Geographic characteristics on square kilometers and density are derived from Census 2000 results. The percent of unmailable addresses are based on the 2000 ACS. Demographic and social characteristics are derived from the 2001 ACS profiles.

The sites are located in 18 different states across the country. They vary in size geographically and demographically, and reflect both urban and rural areas. Their household populations range from 65,357 for the Starr/Zapata, Texas site to 3.7 million for the Fort Bend/Harris, Texas site. The household population density ranged from a low of 6 per square kilometer in Flathead/Lake, Montana to a high of nearly 12,000 per square kilometer in Bronx, New York. We chose the Census 2000 population density as a key factor and ordered the sites from lowest density to highest for tabular and graphic presentation. The ACS relies only on the United States Postal Service (USPS) to deliver questionnaire packages to its sampled addresses. All selected addresses are reviewed for completeness and unmailable addresses are identified, and the extent of unmailable addresses in an area is often reflective of its urban/rural nature. In Census 2000 most addresses received their census questionnaires by mail, but census enumerators delivered forms in some areas (update/leave), and used non-mailback methods in

[3] Thibaudeau, Y. (2002). "Model Explicit Item Imputation for Demographic Categories," Survey Methodology.

[4] Research on how best to decrease the sampling error on these populations by increasing the subsampling rate for personal visit interviewing in areas with very low self-response rates continues, with plans to implement a new subsampling design.

others (list/enumerate, urban or rural update/enumerate). The enumeration method assigned to an area was based on the expected ability of the USPS to deliver mail to specific units and is another measure of the rural nature of an area. Areas with high rates of unmailable or rural-type addresses were likely to have been enumerated using update/leave methods in Census 2000. Table 1 includes the rate of unmailable ACS addresses in each site.

Demographic, social, and economic characteristics also varied across the sites. The percent of non-Hispanic Whites ranges from a low of 4 percent in Starr/Zapata, Texas, to a high of 98 percent in Schuylkill, Pennsylvania. Hispanics account for 30 percent or more of the population in six of the sites (Starr/Zapata, Yakima, Tulare, Pima, Fort Bend/Harris, and Bronx), and Blacks make up 30 percent or more of the population in three sites (Jefferson, Madison, and Bronx). Several sites include high proportions of foreign born (San Francisco, Starr/Zapata, Bronx, and Broward) and many sites include high percentages of persons speaking a language other than English. The proportion of children in poverty varied from under 10 percent in Calvert and Lake to over 30 percent in the Bronx, Jefferson, Tulare, and Starr/Zapata.

Because of their diversity, we were particularly interested in the level of consistency of ACS data quality across the 21 sites. Four basic measures of nonresponse are used in this investigation of ACS and census sample data completeness in small areas, along with a self-response rate. The nonresponse measures are: an overall unit nonresponse rate, a follow-up noninterview rate, a rate of proxy interviews, and summary measures of item nonresponse. The sites are not a probabilistic sample of the nation's 3,141 counties. Most ACS test sites were selected purposively because of their demographics and location, and so the results we describe here cannot be generalized to the nation as a whole. However, within each site we can directly compare results from the 2000 ACS and Census 2000 long forms and subsequent census sample in order to judge their relative levels of data quality. This comparison is especially relevant since the ACS estimates will be replacing the 2010 census sample estimates for these areas. All estimates in this report are based on responses from a sample of the population–the ACS sample and the Census 2000 long form sample. Estimates will vary from actual values because of sampling variation and other factors. All differences noted in this paper are significant at the 90 percent confidence level unless otherwise stated.

The Effect of Environment and Delivery – Levels of Self-Response

Although the extent to which data can be obtained through self-response modes is not a measure of nonsampling error, the success or failure of such a data collection stage determines the size of the follow-up operations that will ultimately decide the quality of the final results. The ACS begins as a **mail-out** survey.[5] Its success from a cost standpoint depends on its ability to mail to samples of addresses and convince the residents of those units to fill out the survey questionnaire and mail it back. The ACS is critically dependent on the quality of the mailing addresses contained on the Census Bureau's Master Address File (MAF). The Census 2000 sample was primarily a **mail-back** survey. Nationally, nearly 80 percent of census housing units were in mail-out areas, but a goodly proportion (19 percent) were in areas where questionnaires were not mailed out through the USPS but delivered to the units by census update/leave enumerators for mail back by respondents. Census 2000 also used non-mail direct enumeration methods for about 1 percent of units nationally. Two of the 21 ACS sites involved in this analysis – those having the lowest household population densities– had significant proportions of their units enumerated in the census using such methods.

We are using a weighted self-response rate to compare the effectiveness of self-response between the 2000 ACS and Census 2000. For the ACS, it reflects the percent of all occupied units (households) in the site that responded to the survey by mailing back the questionnaire. In areas with high rates of unmailable addresses, this rate is expected to be low. For Census 2000 it reflects the percent of occupied units (households) enumerated on long form questionnaires that were mail returns. The census rate is expected to be low in areas where non-mailback methods were also used. The self-response rate for both the ACS and for the Census 2000 long form is a weighted measure designed to address the fact that an area's sampling rate can have an effect on its level of self-response. This is particularly important if the measure is used to compare self-response between census areas sampled at different rates, such as those sampled at 1-in-2 versus those sampled at 1-in-8. The ACS has adopted the census's differential sampling scheme, which means that areas within each of the 21 sites in this analysis may also have been sampled at different rates. A consistent self-response measure was computed for both the census and the ACS by weighting the units in the numerators and the denominators by their initial probabilities of selection, and for ACS personal visit units, by their subsampling factors as well.

The self-response rates are shown in Figure 1. The sites are ordered by density from left to right, the least

[5] The 2000 ACS's ability to mail was dependent on the mailability of the addresses it selected from the Master Address File (MAF). If the sample addresses were deficient the mail-out could not take place. Two-thirds of the unmailables were systematically selected to be interviewed by personal visit followup. The denominator for the ACS self-response rate includes all housing units regardless of whether they were actually mailed to.

dense shown left-most on the X axis. All site-level graphs and tables in this report display the 21 sites in this same density order. The Census 2000 long form self-response rates are higher than the ACS rate in all but one site. This was not unexpected. The effect of the decennial census environment alone on public cooperation has been found to increase mail return rates over those observed in census tests by 10 to 20 percentage points[6]. The self-response rates in 15 of the 21 ACS sites differed by 10 percentage points or less with the census long form rate, and all were within 19 percentage points. Census 2000 enumerated about one-fifth of the housing units in Flathead/Lake and two-fifths of the units in Starr/Zapata, the two most sparsely populated of the ACS sites, by non-mailback methods. The condition of the addresses in the Starr/Zapata site hampered the level of ACS self-response and probably influenced the decision to use conventional direct enumeration methods in that area for the census, while it appears that the addresses in the Flathead/Lake site were better than expected, and resulted in a successful ACS mail data collection.

The Effect of Followup Operations – Levels of Noninterview

The level of self-response is important for both the census sample and the ACS since it reflects the relative amount of information collected by operations usually referred to as nonresponse follow-ups. From a data quality standpoint, the consistently lower ACS self-response rates in these sites meant that the overall success of the ACS depended heavily on the survey's followup operations. Successfully following up with the households who do not choose to complete and return their mail questionnaires will determine the quality of the final sample estimates. The accuracy of the final data distributions for areas with low self-response will depend heavily on the thoroughness of the follow-up. The households that completed and mailed back their ACS questionnaires and census long forms are a self-selected group whose characteristics are often very different from the rest of the sampled households whose cooperation was harder to win. The extent to which the ACS and census had difficulty in obtaining a "successful" response – an "acceptable" interview in the ACS and a "sample data-defined" long form in the census – affected the accuracy of the final estimates.

Noninterview rates for the Census 2000 sample and for ACS were calculated separately for self-response and follow-up sample units for each of the 21 sites, weighting the units involved in each mode by their probabilities of selection as we did in computation of the self-response rates. The universe includes both vacant and occupied housing units. The Census 2000 follow-up mode consists of all units enumerated on long forms that were not mail returns. Similarly, the ACS follow-up mode consists of all sample units not interviewed by self-response modes.[7] National studies of ACS and Census 2000 long form mail returns have shown that nearly all meet the minimum data requirements of their respective processes to be considered interviews.[8] Our 21 sites were no different. About 99 percent of all Census 2000 housing units enumerated on long form mail return questionnaires were placed in the census samples of the 21 sites, and about 98 percent of all 2000 ACS units interviewed through self-response means were treated as interviews. The vast majority of the unsuccessful long form enumerations and ACS noninterviews are from the followup operations conducted with the initially uncooperative households.

The overall impact can be seen in Figure 2. The 2000 ACS followup noninterview rate is considerably lower than the Census 2000 long form followup rate in every site. What is particularly noteworthy is that the four sites showing the greatest level of difficulty collecting Census 2000 sample data in follow-up operations are also the sites that exhibit the largest differences with ACS. The ACS's level of successful followup appears to be more consistent across the different sites, with the lowest density sites showing some of the lowest levels of noninterview. The simple average of the census follow-up noninterview site rates is 21.5 percent, with individual site rates deviating from this average by an average 7.4 percentage points. The comparable measures for the ACS are a site average of 6.9 percent with an average deviation of 2.5 percentage points among the sites. The 2000 ACS followup operations had noninterview rates that were more than 25 percentage points lower than the comparable Census long form rates for Jefferson, Madison, Black Hawk, and the Bronx. The Census 2000 rates in this graph have a direct effect on the quality of the census sample data for these sites since they reflect the percent of the all enumerator-filled long form questionnaires that were not included in the Census 2000 samples for these areas.

An additional indicator of follow-up data accuracy is the proxy rate. In the parlance of the decennial census, a proxy rate measures the extent to which information is collected from people who are not members of the

[6] "Influence of 13 Design Factors on Completion Rates to Decennial Census Questionnaires," Don A. Dillman, Jon R. Clark, and James B. Treat, April 5, 1994, presented at the 1994 Annual Research Conference;
2KS Memorandum Series, Design 2000, Book 1, Chapter 14, #24, May 29, 1992, "Mail Response/Return Rates by Type of Form – 1970, 1980, and 1990".

[7] Self-response in ACS consists of mail returns and interviews completed by the Telephone Questionnaire Assistance staff, while long form self-response in Census 2000 consists only of mail returns.

[8] Data on national long form sample data defined rates by mode presented as a supplement to March 11, 2003 presentation at the Census Bureau, "Making it in the Bronx," by Joseph Salvo.

sample household. It is important to mention this because the ACS data collection procedures for both the telephone and the personal visit computer-assisted follow-up operations prohibit interviews with occupied units from being conducted with anyone other than a member of the ACS sample household. In contrast, the census procedures allow neighbors or others to provide information about households not their own after attempts to contact the household members have failed. Data collected from proxy respondents have repeatedly been shown to be less accurate, the most recent assessment of Census 2000 coverage bearing this out, at least for the basic demographic data collected for all households.[9] Relying on proxy information would seem to be even more problematic when it involves the collection of data used in detailed demographic and socio-economic estimates of the kind produced by the census samples.

Figure 3 illustrates the proxy rate in each of the 21 sites for occupied units enumerated on long forms filled by census enumerators. The simple average proxy rate for all occupied long forms completed during follow-up in the 21 ACS sites was 19.3 percent. It is not unreasonable to consider the combination of proxies and noninterviews as a measure of the success of follow-up operations.

The Success of Data Collection – Levels of Unit Nonresponse

Unit nonresponse is the most commonly used measure of overall data collection success in a survey, and an important indicator of potential nonsampling error in the final estimates. To compare the level of unit nonresponse in the ACS and the Census 2000 sample, we again defined similar measures for both data sources. A weighted unit nonresponse rate is computed for the ACS on a yearly basis as part of the Census Bureau's required survey data quality profile. Units in the ACS sample that are not successfully interviewed are weighted by their probabilities of selection (and subsampling factors, if applicable), summed, and expressed as a percent of the final total sample units similarly weighted. To be considered successfully interviewed a unit's data have to meet or exceed an acceptability or data completeness standard based on the completeness of the basic population items for each household. Failing households are considered noninterviews and comprise the numerator of the unit nonresponse rate. Vacant units are not subjected to the minimum data requirement but are included in the denominator.

We defined a comparable measure of unit nonresponse for the Census 2000 sample using its data completeness standard. The Census Bureau has been using a minimum data criteria for occupied long form questionnaires at least since the 1970 census to prevent households for which no sample population data were collected from being placed in the census sample. A similar standard for vacant units was introduced for Census 2000. These "non-sample data defined" units in each of the 21 sites represent most of the census sample noninterviews in our Census 2000 unit nonresponse measure. The remaining component of census sample unit nonresponse is the result of insufficient numbers of units being enumerated on long forms regardless of their data-defined status. To arrive at the overall measure of unit nonresponse rate for the Census 2000 sample in each of the 21 sites, we weighted the long form units that met the minimum data criteria for inclusion in the sample by their probabilities of selection, summed them, and subtracted this weighted total from the full ("100%") census count of units in the site. This difference was then divided by site's full census count of units. The numerator of this rate represents the shortage of housing units in the Census 2000 sample, and the denominator represents the total housing being represented by the census sample.

The ACS noninterviews remain as members of the ACS sample and are adjusted for in the survey's weighting and estimation process by noninterview factors. The census sample non-data defined units do not remain as members of the census sample. Units that are placed in the census sample are weighted to the full census count of housing units and population by demographic cohorts based on the "100%" characteristics. The potential for the introduction of nonsampling error and bias into sample estimates is directly related not only to the overall level of unit nonresponse but to the extent to which units and their occupants not included in the sample differ from those that are.

The unit nonresponse rates described above are compared in Figure 4. The ACS unit nonresponse rates are lower than the Census 2000 sample rates in 20 of the 21 sites, and in most there is a considerable difference between the two.[10] In addition, the ACS levels of unit nonresponse appear to be more consistent across all densities than the Census 2000 sample nonresponse rates, with sites with the lowest densities (on the far left) faring at least as well as the higher density sites (on the far right). All of the 21 sites had census sample unit nonresponse rates of 5 percent or more, while only three ACS sites were above that level– the Bronx, San Francisco, and Pima. But the census level of sample nonresponse in these three sites was twice the ACS level. The simple average of the census measures for the sites was 9.7 percent, with an average deviation of 3.3 percentage points between sites, while the comparable average ACS site measure was 4.0 percent with a deviation of only 1.1 percentage points. The complement of these total unit nonresponse

[9] Technical Assessment of A.C.E. Revision II, March 12, 2003

[10] The unit nonresponse rates for Census 2000 and ACS in the Sevier, Tennessee site were not significantly different.

rates are the overall weighted survey response rates. Given the scale of the decennial data collection, the Census 2000 sample response rates are quite respectable. The priority placed on obtaining the population count and the tight time frame in which to collect enormous amounts of information from a growing and more diverse population, has in recent censuses resulted in increasing difficulty collecting the required long form sample data. The success of the ACS data collection operations is evident in every one of these 21 sites that vary not only in density but in demographic, social, and economic characteristics. Sites with less than 50 percent of the population reporting as White, nonHispanic, fared much better in the ACS. Sites with high rates of nonEnglish speakers also did significantly better in the ACS. ACS response appears to be considerably more consistent across these very different test sites than was response to the Census 2000 sample enumeration.

Completeness of the Final Data – Levels of Item Nonresponse

Unit nonresponse is only one measure of potential nonsampling error. It impacts the quality of all estimates produced by a sample depending on the size and nature of the nonresponse, and the success of any weighting noninterview adjustments devised to correct for it. Item nonresponse–the extent to which a required answer to an individual questionnaire item is missing–can be an even greater threat to the quality of the final survey distributions. Getting complete answers is much more difficult for some questions than for others. Some are considered too personal by many respondents, some require that household records be reviewed, while some questions ask for information that is just not known by the respondent. Self-response modes add another complication – the ability of respondents to wend their way through the various paths and patterns of a questionnaire on their own. Collecting complete information from sample units that have been unresponsive to prior requests to complete and return a mail-back questionnaire can be very difficult, but mail nonresponse universes often consist of the very units and people whose characteristics are critical for the production of the survey's most important estimates.

To compare item nonresponse levels between Census 2000 sample data and the 2000 ACS data, we restricted the items involved to those that the two data collection operations had in common, and within that group, to the items that were asked of occupied housing units and their residents and whose edit output was comparable. Fifty-four population items and 29 housing items were included in the comparison. The measure commonly used to reflect the level of nonresponse for an edited item is the item's final weighted allocation rate–comparable to the rate that can be calculated from the allocation tables released with both the ACS and the census estimates. The allocation rate for an item is the percent of the item values imputed by the edit from a different housing or person record because answers are missing or inconsistent. To summarize the results and look at overall data completeness, we computed a summary allocation rate for all items together. These rates provide a measure across all data items for occupied units, and are derived from the ratio of all household population and occupied housing unit items for which the content edits allocated values to the total number of household population and occupied housing items required to have a response. These composite measures provide a summary picture of the completeness of all data. These rates are shown in Figure 5 for each site.

The overall item nonresponse measure for the 2000 ACS data is lower in every site than the comparable Census 2000 item nonresponse measure. In some sites the difference is considerable. Twelve of the sites had Census 2000 summary allocation rates of 10 percent or more. Areas with the lowest population densities were as likely as high density areas to have these high rates. The level of allocation in the 2000 ACS data deviates less across sites than the level in the Census 2000 sample. The simple average of the site summary allocation measure in the census is 10.6 percent with an average deviation between sites of 1.3 percentage points, while the average ACS summary allocation is 6.6 percent with a deviation of 0.8 percentage points.

To interpret these allocation rates and their potential to add nonsampling error to the estimates, we need to consider the level of unit nonresponse and item nonresponse together. The only information available to the content edits in either the census or the ACS is the data on the response records considered interviews. The noninterviews in the Census 2000 sample were not even placed on the sample file. The "nearest neighbor" hot deck matrices used to allocate in both the census and the ACS sample process use item values reported on one interview record to fill in missing information on another interview record. If the characteristics of the noninterview sample units differ importantly from those of the interview sample units, and the interview records missing considerable amounts of information get their imputed values primarily from interview records with very little missing data, it is reasonable to assume that high unit nonresponse combined with high item nonresponse can have a detrimental affect on the accuracy of the data.

Conclusions

The decennial sample has, particularly since the 1960 census, produced important detailed housing, demographic, and socio-economic estimates of population and housing characteristics down to at least the level of census tracts. More recently the census sample has become increasingly important as the number of questions asked of everyone (the "100%" items) have been limited to six. The funding of most federal programs depends on the quality and completeness of the information available from sample surveys, with the decennial census sample estimates stipulated in legislation as the source required for administering many governmental programs. It has become more apparent that mounting the huge data collection effort once every ten years to produce the detailed estimates needed by all levels of government, private industry, research organizations, and even the general public is no longer viable. Not only are the results from the census sample obsolete before they become available, but the enormity and cost of collecting "long form" data while also conducting a complete count of the nation's population and housing is becoming more obvious to all involved.

The ACS has been specifically designed to replace the decennial census sample design, data collection, processing, estimation, and distribution, leaving the census to concentrate all its efforts on the quality of the once-a-decade full census count required by the Constitution, and the collection of the basic characteristics commonly known as the "short form" items that are essential for this count. The ACS will replace the census sample by dividing the sample into monthly pieces and the data collection into manageable assignments for a highly-trained permanent interviewing staff using computer-assisted survey instruments. Data collected from these continuous samples will be aggregated for all areas over five year periods, with single-year distributions released for areas with populations of at least 65,000 people, three- year average distributions released for areas of 20,000 or more, and five-year average distributions released for every tract and block group based on five years of aggregated sample data. Once five years of data have been accumulated, the aggregate will be refreshed every year by dropping the data from the earliest year and adding the most current yearly data.

This paper has presented a first-hand look at data produced by the 2000 ACS program in a group of counties purposively selected to participate in the ACS testing program, a program that began producing distributions of estimates in 1996. We used several measures of nonsampling error related to issues of nonresponse to gauge the quality and completeness of the ACS data and derived comparable nonresponse measures from Census 2000 long form sample results. The 21 ACS test sites used for this analysis represented a wide range of counties from the high density and highly diverse borough of the Bronx in New York City to the sparsely populated ranges of Flathead and the American Indian reservations of Lake in Montana. The results are clear. The methods that the ACS used to collect the eclectic content previously available only through administering the "long form" to a sample of housing units during the conduct of the decennial census resulted in estimates in each ACS site with considerably lower levels of both unit and item nonresponse than those produced from the census sample. Analysis of ACS test site results from earlier years have shown similar low levels of nonresponse when compared to the 1990 census sample results, while an initial review of post-2000 ACS results has found nonresponse levels to be the same or lower than those in this study. Sampling error on the estimates produced by decennial census samples for small areas may in fact be lower than on estimates from the ACS, but the levels of nonsampling error on the census sample estimates may negate the difference. A study based on three years of ACS data for all 36 ACS counties is being conducted that will investigate the differences in the distributions produced by the Census 2000 sample and the ACS by comparing the actual distributions of the population and housing characteristics produced by both programs, their sampling errors, and the same nonsampling error measures of unit and item nonresponse that we have used in this site analysis, but at sub-county levels. We look forward to those results.

Table 1. Selected Geographic and Demographic Characteristics of the 21 ACS Test Sites

ACS Test Site	**Square Kilo-meters**	**Census 2000 Household Population**	**Density**	**Percent unmailable addresses**	**Percent White, Non-Hispanic**	**Percent Foreign Born**	**Percent language other than English**	**Percent of children in Poverty**
Flathead/Lake MT	17074	99217	6	16	89	2	4	17
Starr/Zapata TX	5750	65357	11	61	4	31	95	59
Yakima WA	11127	218844	20	0	55	17	34	29
Tulare CA	12495	361980	29	2	41	22	44	33
Jefferson AR	2292	78989	34	3	47	1	2	31
Pima AZ	23794	821712	35	1	60	11	27	21
Madison MS	1863	72615	39	1	60	2	3	15
Sevier TN	1534	70533	46	10	96	2	3	19
Schuylkill PA	2017	143110	71	4	98	1	4	17
Black Hawk IA	1470	121535	83	0	87	4	6	17
Calvert MD	557	73982	133	5	82	3	5	3
Hampden MA	1602	441799	276	1	74	8	22	21
Broward FL	3131	1603094	512	0	56	26	29	15
Douglas NE	857	451878	527	0	78	7	10	12
Lake IL	1160	623378	538	0	73	16	23	8
Ft. Bend/Harris TX	6744	3706598	550	1	41	22	36	18
Multnomah OR	1127	643798	571	0	76	13	17	17
Rockland NY	451	279104	619	0	71	20	29	12
Franklin OH	1399	1046872	749	0	74	7	9	15
San Francisco CA	121	756976	6258	0	43	38	46	11
Bronx NY	109	1285415	11793	0	14	30	55	41

Figure 1: Comparison of Self-Response Rates for the Census 2000 Long Form and the 2000 ACS, based on Occupied Housing Units weighted by the probabilities of selection.

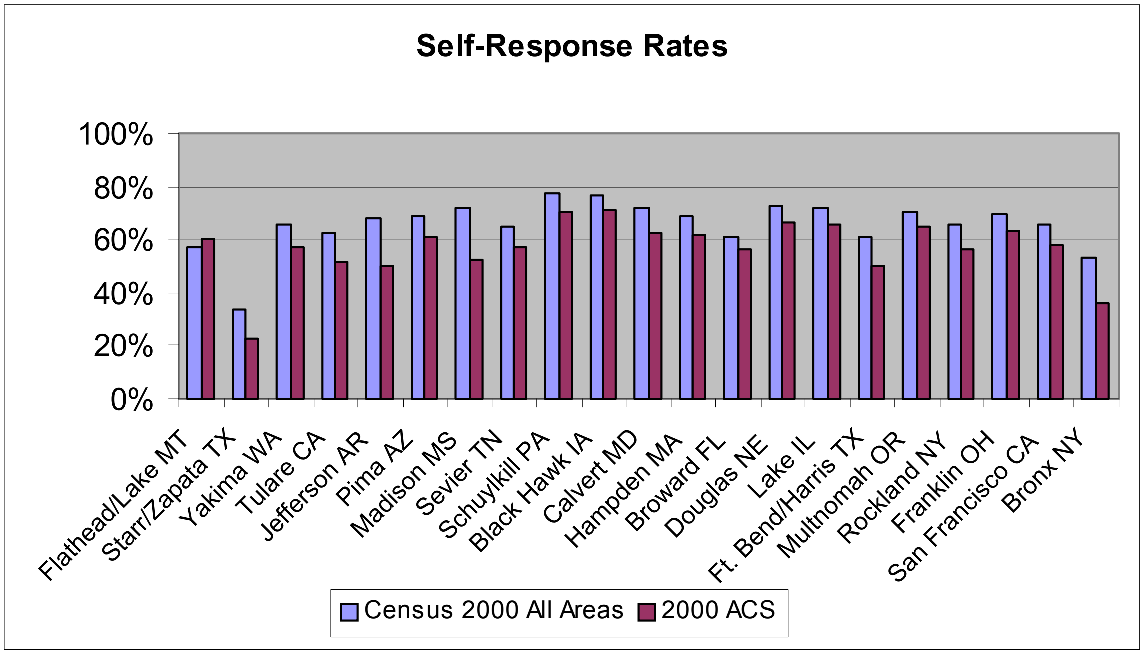

Figure 2: Comparison of the Level of Noninterview in Follow-up Operations for Census 2000 Long Forms and the 2000 ACS, based on Occupied and Vacant Housing Units weighted by the probabilities of selection.

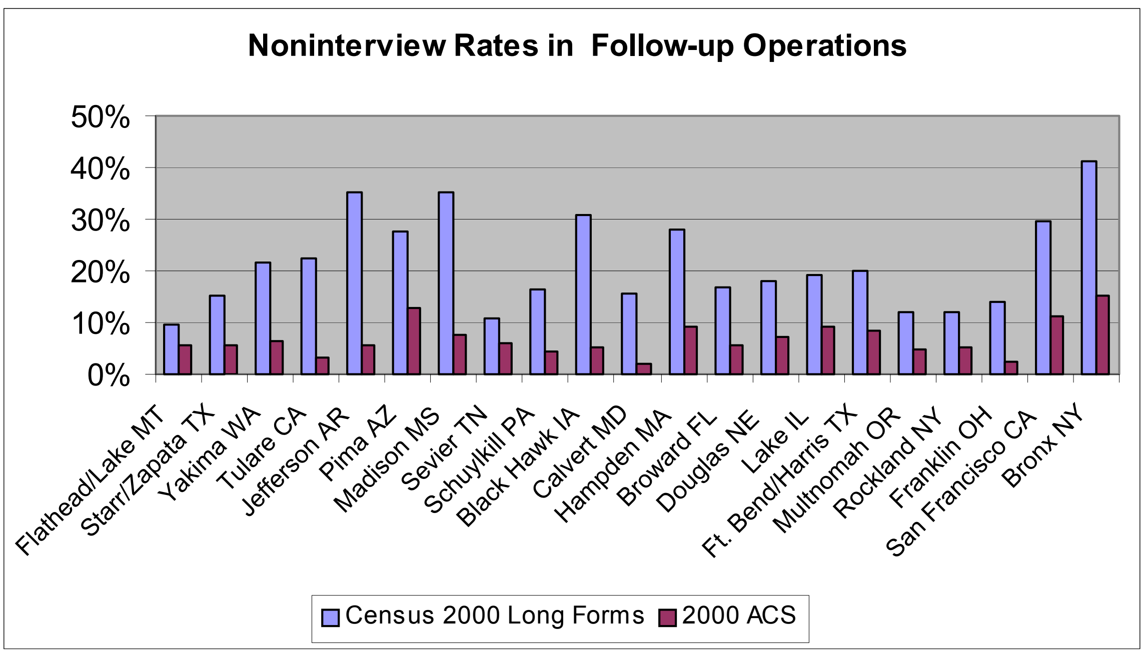

Figure 3: The Level of Proxy Enumeration of Occupied Housing Units for Enumerator-filled Long Forms in Census 2000

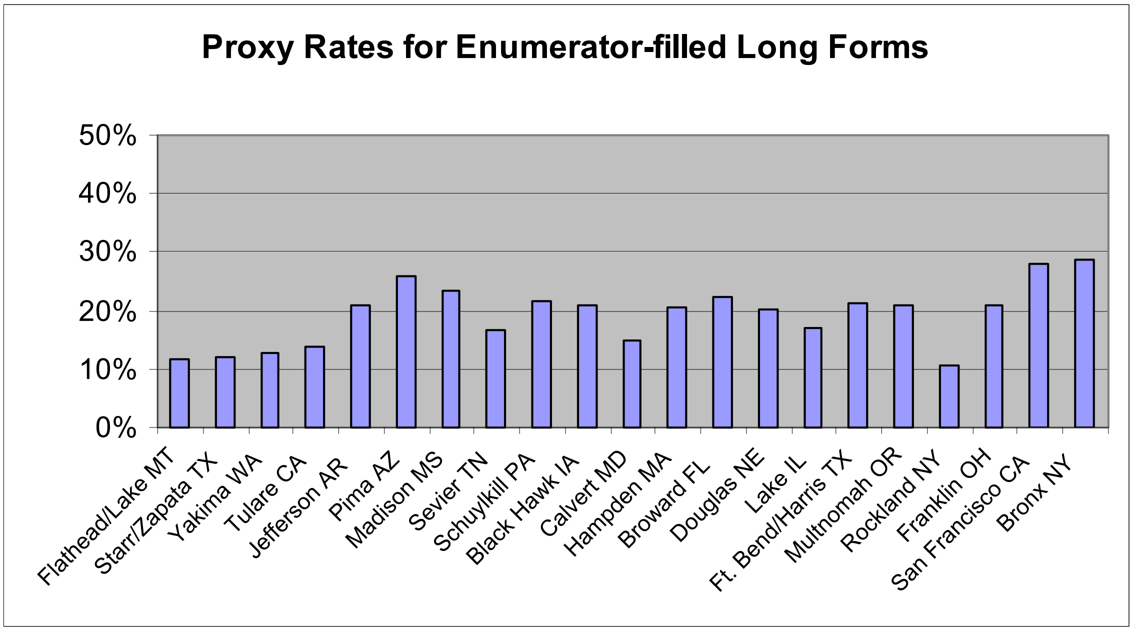

Figure 4: Comparison of Total Unit Nonresponse Rates for the Census 2000 Sample and the 2000 ACS, weighted by the probabilities of selection.

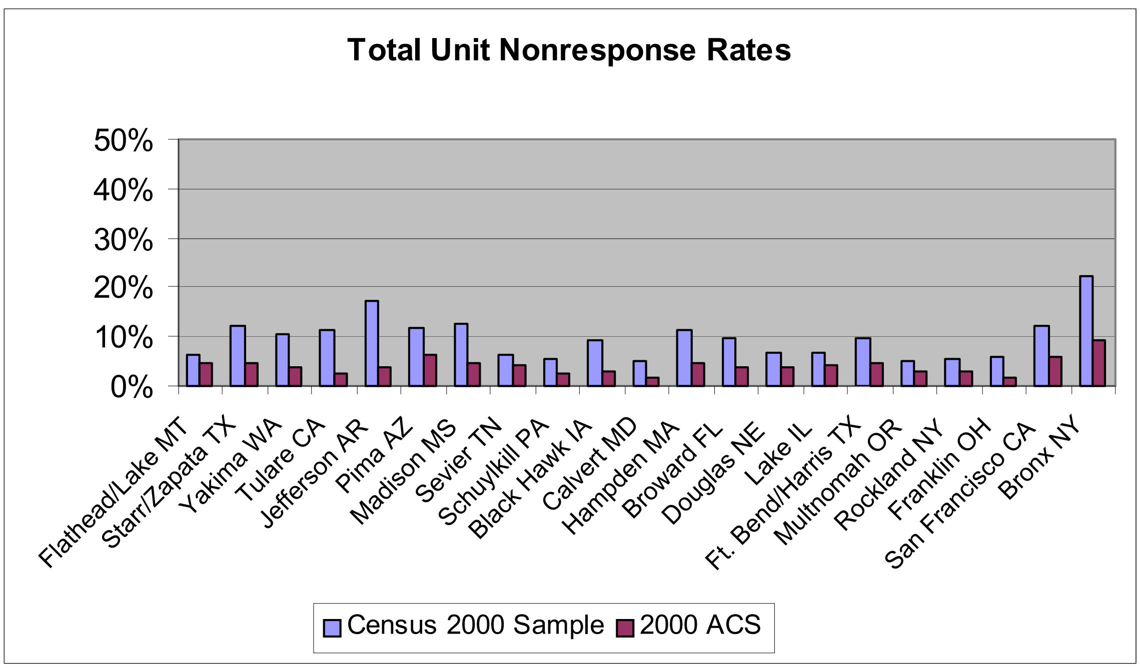

Figure 5: Comparison of Summary Item Allocation Rates in the Census 2000 Sample and 2000 ACS for Occupied Housing Units, based on final weighted data.